interested in Porsche Boxster cars. Indeed, there are several explicitly-gathered resource collections, such as those listed under the category of "Recreation: Automotive: Makes and Models: Porsche: Boxster" at the Yahoo Web site (yahoo.com), which are devoted to the Boxster. Most of these communities manifest themselves as news groups, Web rings, or as resource collections in directories such as Yahoo! and Infoseek, and as homesteads on Geocities. Other examples include popular topics such as "Major League Baseball," or the somewhat less visible community of "Prepaid phone card collectors". The explicit nature of these communities makes them easy to find. It is simply a matter of visiting the appropriate portal or news groups.

On page 8 please delete the second full paragraph under the <u>Strongly-connected bipartite subgraphs and cores</u> section and replace with the following replacement paragraph:

Linkage between the related pages can nevertheless be established by a different phenomenon that one can observe on the Web: pages focusing on the same theme frequently contain hyperlinks to the same pages. For instance, as of 12/1/98 the sites swim.org/church.html, kcm.co.kr/search/html, korea.html, and cyberkorean.com/church all contain links to numerous Korean churches. This phenomenon is referred to as co-citation, which originated in the bibliometrics literature. See, for instance, Bibliometrics, Annual Review of Information Science and Technology, volume 24, pages 119-186, Elsevier, Amsterdam, 1989. Co-citation suggests that related pages are frequently referenced together. This is even more true in the Web world where linking is not only indicative of good academic discourse, but the essential element that distinguishes the Web as a corpus from other text corpora. For example, the corporate home pages of AT&T and Sprint typically do not reference each other. On the other hand, these pages are very frequently "co-cited". Co-citation is not just a characteristic of

well-developed and explicitly-known communities (such as the ones listed above) but an early indicator of newly emerging communities. In other words, the structure of such co-citation in the Web graph can be exploited to extract all communities that have taken shape on the Web, even before the participants have realized that they have formed a community through their co-citation.

On page 11-12 please delete the first paragraph of the "(c) In-degree distribution" section and replace with the following replacement paragraph:

The first approach to trimming down the resulting data came from an analysis of the in-degrees of Web pages. The distribution of page in-degrees has a remarkably simple rule, as can be seen in the chart of FIG. 4 This chart includes pages that have in-degree at most 410. For any integer k larger than 410, the chance that a page has in-degree k is less than 1 in a million. These unusually popular pages (e.g., yahoo.com) with many potential fans pointing to them have been excluded. The chart suggests a simple relation between in-degree values and their probability densities. Indeed, as can be seen from the remarkably linear log-log plot, the slope of the curve is close to ½. This leads to the following empirical fact: the probability that a page has in-degree i is roughly $1/i^2$.

On page 13 please delete paragraph one of the section entitled "Trawling" and replace with the following replacement paragraph:

Thus far, several preliminary processing steps on the data have been described, along with some interesting phenomena on degree distributions on the Web graph. The trawling of this "cleaned up" data for communities is now described in detail. The test data still has over 2 million potential fans remaining, with over 60 million links to over 20 million potential centers. Since there are still several million potential fans, it is not practical to enumerate the communities in the form "for all subsets of i potential fans, and for all subsets of j potential centers, check if a core is induced". A number of additional pruning steps are therefore necessary to eliminate much of this data, while retaining the property that the

3

eliminated nodes and links cannot be part of any core that is not explicitly identified and output before they are pruned. After the data is reduced by another order of magnitude in this fashion, enumeration of the communities may begin.

On page 18-19 please delete the second paragraphs of the section entitled "c) Core generation and filtering" and replace with the following replacement paragraph:

Next, nepotistic cores are removed. A nepotistic core is one where some of the fans in the core come from the same Web site. The underlying principle is that if many of the fans in a core come from the same Web site, this may be an artificially established community serving the ends (very likely commercial) of a single entity, rather than a spontaneously-emerging Web community. For this purpose, the following definition of "same Web site" is used. If the site contains at most three fields, for instance, yahoo.com, or ibm.com then the site is left as is. If the site has more than three fields, as in yahoo.co.uk, then the first field is dropped. The last column of Table 1 represents the number of non-nepotistic cores. As can be seen, the number of nepotistic cores is significant, but not overwhelming. About half the cores pass the nepotism test.

On page 21 please delete the section entitled "Communities" and replace with the following replacement paragraph:

Next, the communities themselves were studied. The following two examples give a sense of the communities that were identified. The first one deals with Japanese pop singer Hekiru Shiina, which has the following fans:

awa.a-web.co.jp/~buglin/shiina/link.html

hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html

noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html

The next example deals with Australian fire brigade services with the following fans:

maya.eagles.bbs.net.au/~mp/aussie.html

4

homepage.midusa.net/~timcorny/intrnatl.html

fsinfo.cs.uni-sb.de/~pahu/links_australien.html